

d) Tabelas de resultados – exemplo.txt

O resultado esperado do *dadosSemiotica* não é a conclusão de sua monografia, artigo ou tese: é uma tabela. Para que uma tabela?

A tabela com os resultados é o que permite:

- visualização de resultados de forma organizada
- análises cruzadas de categorias
- análises estatísticas
- inclusão de dados das análises automáticas feitas pelo sistema (NA VERSÃO 1.x, trata-se de resultados das análises morfossintáticas, por meio de expressões tregex)

É assim que o pesquisador pode chegar a conclusões consistentes sobre o corpus a partir das análises feitas.

O *dadosSemiotica* possui um módulo de pós-processamento estatístico que é acessado pelo ícone de gráfico ao lado do nome do projeto (figura 9) e:

- retorna as tabelas com os dados desejados (sempre do conjunto total de textos do projeto, mas é possível especificar quais as categorias que deseja visualizar na tabela)
- permite escolher se deseja ou não incluir na tabela as sentenças e o conteúdo do *corpus* (o conteúdo do corpus são os parágrafos, o que, no caso desse exemplo, é um pedaço da sentença original).
- Permite acessar os resultados da análise morfossintática pela inclusão de buscas TRegex (NA VERSÃO 1.0, a opção “montar expressão simples” ainda não está disponível e é necessário criar a expressão usando o padrão da página indicada no próprio *dadosSemiotica*; na versão 1.5 está disponível mas ainda incompleta).

Ao abrir o módulo de pósprocessamento estatístico, temos acesso à lista de categorias do projeto *ds* (figura 20).

dadosSemiotica: projetos ds

Marque as categorias que deseja incluir no arquivo CSV e depois marque aquelas que pretende gerar o histograma.

S.	Categoria	H.	Título do Histograma
<input type="checkbox"/>	Texto-partes	<input type="checkbox"/>	Texto-partes
<input type="checkbox"/>	Texto-comentários-geral	<input type="checkbox"/>	Texto-comentários-geral
<input type="checkbox"/>	Texto-comentários-lembretes	<input type="checkbox"/>	Texto-comentários-lembretes
<input checked="" type="checkbox"/>	Semiótica-categoriaProfunda	<input checked="" type="checkbox"/>	Semiótica-categoriaProfunda
<input checked="" type="checkbox"/>	Semiótica-categoriaProfunda-foria	<input type="checkbox"/>	Semiótica-categoriaProfunda-foria
<input checked="" type="checkbox"/>	Semiótica-categoriaNarrativa-S1	<input type="checkbox"/>	Semiótica-categoriaNarrativa-S1
<input checked="" type="checkbox"/>	Semiótica-categoriaNarrativa-S2	<input type="checkbox"/>	Semiótica-categoriaNarrativa-S2
<input checked="" type="checkbox"/>	Semiótica-categoriaNarrativa-O	<input type="checkbox"/>	Semiótica-categoriaNarrativa-O
<input checked="" type="checkbox"/>	Semiótica-categoriaNarrativa-estado	<input type="checkbox"/>	Semiótica-categoriaNarrativa-estado

Ilustração 20: topo da página para gerar as estatísticas; o nome da categoria é sensível e é possível marcar/desmarcar a categoria clicando sobre ela.

É possível obter resultados mesmo antes de ter todas as categorias analisadas. Por exemplo, é interessante, após preencher as categorias do grupo Texto, obter uma tabela completa com as anotações para procurar por algum padrão que indique a relevância de alguma análise não prevista inicialmente no projeto.

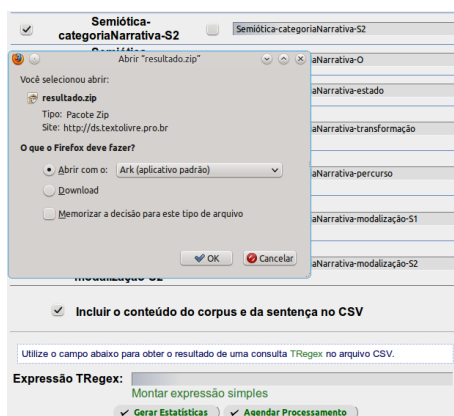


Ilustração 21: Popup para download dos resultados.

O resultado obtido é sempre um arquivo comprimido contendo os arquivos solicitados (figura 21).

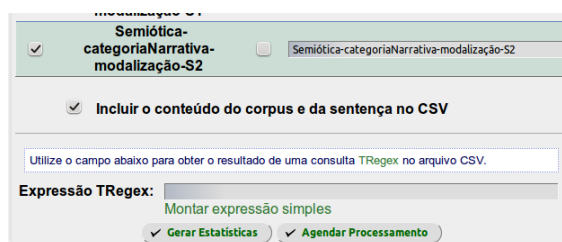


Ilustração 22: Para gerar as tabelas, é possível escolher entre "Gerar Estatísticas" ou "Agendar processamento". A segunda opção é útil para corpus de grandes dimensões, pois não é preciso ficar online aguardando o resultado.

No caso dessa primeira seleção que fizemos (figura 22), e solicitando o histograma da categoria profunda (figura 20), o resultado obtido compreende dois arquivos e um diretório:

- Arquivo csv: no caso do *exemplo.txt*, chama-se *projeto2.csv* e contém a tabela com as categorias requisitadas. O número corresponde ao número do projeto no banco de dados.
- Arquivo debug.txt: contém informações que podem ser úteis caso algo venha a sair errado. Guarde-o apenas para fins de suporte.
- Diretório categoria1: contém os arquivos relativos à categoria cujo histograma foi solicitado (figura 24):
 - Arquivo categoria.csv: contém a tabela com os dados exclusivos da categoria para a qual foi solicitado o histograma.
 - Arquivo descritivo.txt: contém resultados de uma análise estatística simples da categoria em questão: média, mediana, variância, desvio padrão, mínimo e máximo. No exemplo (copiado abaixo), está registrado que a *variável 1* corresponde à categoria (*nome*) *Texto.Partes.n¹*, possui 15 *observações* (dados não nulos), nenhum *valor perdido* (nulo), *média* igual a 2, *variância* igual a 2.142857, *desvio padrão (SD)* igual a 1.46385, *valor mínimo* igual a 1, *máximo* igual a 4 e *mediana* igual a 2. A análise desses valores me permite dizer, dentre outras coisas, que não houve nenhum erro de digitação nas análises e todas as sentenças foram analisadas.

Variable Name	Number of obs	Number of missing values	Mean	Variance	Sd
1 Texto.Partes.n	15	0	2	2.142857	1.46385
Minimum Maximum Median					
1	0	4	2		

- Arquivo histograma.jpg: contém uma imagem com o gráfico do histograma da categoria em foco².

O diretório só é criado se um histograma for solicitado e haverá um diretório para cada histograma.

1 o .n está presente no nome da categoria porque todas as categorias são passadas para numéricas para permitir a criação do histograma e desses resultados.

2 O histograma é um gráfico que representa o número de amostras para cada resultado de análise. É interessante em categorias fechadas. No caso do Texto-Partes, pode mostrar se o tamanho das partes é igualmente distribuído no corpus ou alguma parte possui mais sentenças que outras.

dadosSemiotica: projetos ds

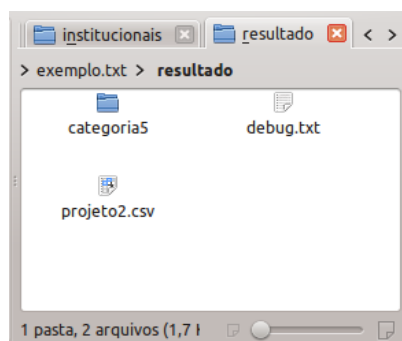


Ilustração 23: Arquivos após descompactação do resultado.zip.

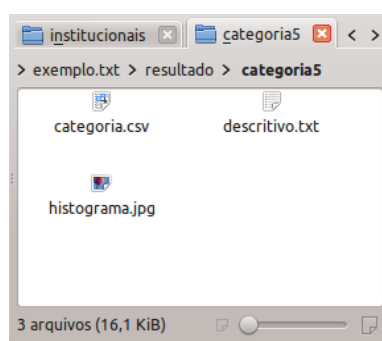


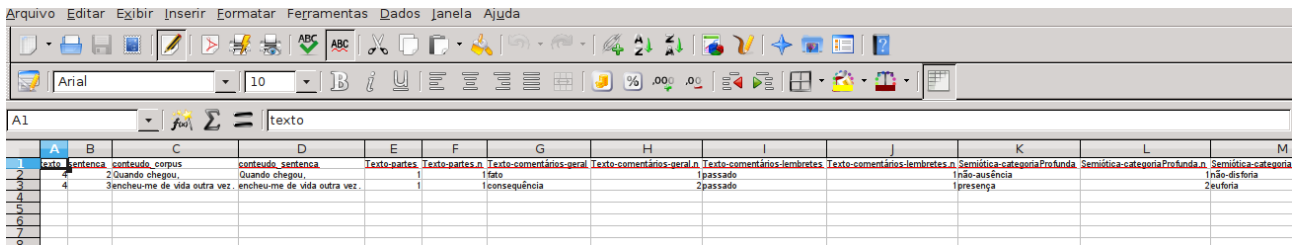
Ilustração 24: Arquivos do diretório categoria5.

O arquivo csv pode ser aberto num editor de texto:

```
texto;sentenca;conteudo_corpus;conteudo_sentenca;Texto-partes;Texto-partes.n;Texto-comentários-geral;Texto-comentários-geral.n;Texto-comentários-lembreres;Texto-comentários-lembreres.n;Semiótica-categoriaProfunda;Semiótica-categoriaProfunda.n;Semiótica-categoriaProfunda-foria;Semiótica-categoriaProfunda-foria.n;Semiótica-categoriaNarrativa-S1;Semiótica-categoriaNarrativa-S1.n;Semiótica-categoriaNarrativa-S2;Semiótica-categoriaNarrativa-S2.n;Semiótica-categoriaNarrativa-O;Semiótica-categoriaNarrativa-O.n;Semiótica-categoriaNarrativa-estado;Semiótica-categoriaNarrativa-estado.n;Semiótica-categoriaNarrativa-transformação;Semiótica-categoriaNarrativa-transformação.n;Semiótica-categoriaNarrativa-percurso;Semiótica-categoriaNarrativa-percurso.n;Semiótica-categoriaNarrativa-modalização-S1;Semiótica-categoriaNarrativa-modalização-S1.n;Semiótica-categoriaNarrativa-modalização-S2;Semiótica-categoriaNarrativa-modalização-S2.n
4;2;"Quando chegou,;"Quando chegou,;"1;1;fato;1;passado;1;não-ausência;1;não-disforia;1;NA;NA;ele;1;NA;NA;não-disjunção;1;NA;NA;sanção;1;NA;NA;atualizado;1
4;3;"encheu-me de vida outra vez.;"encheu-me de vida outra vez.;"1;1;consequência;2;passado;1;presença;2;euforia;2;eu;1;ele;1;vida;1;conjunção;2;conjunção;1;sanção;1;virtualizado;1;realizado;2
```

Se aberto como planilha (figura 25), numa suíte de escritório como o openoffice, terá a seguinte aparência (o mesmo arquivo projeto2.csv acima):

dadosSemiotica: projetos ds



A	B	C	D	E	F	G	H	I	J	K	L	M
1	texto_sentenca	conteudo_corpus	conteudo_sentenca	Texto-partes	Texto-partes.n	Texto-comentarios-geral	Texto-comentarios-geral.n	Texto-comentarios-lembretes	Texto-comentarios-lembretes.n	Semiotica-categoriaProfunda	Semiotica-categoriaProfunda.n	Semiotica-categoria
2	4	2 Quando chegou,	Quando chegou,	1	1rato			1passado		1nao-ausencia		1nao-distonia
3	4	3 encheu-me de vida outra vez,	encheu-me de vida outra vez,		1consequencia			2passado		1presenca		2evofonia
4												
5												
6												
7												
8												

Ilustração 25: Tabela do exemplo.txt aberta no LibreOffice.

Os campos nulos são preenchidos com “NA”.

O arquivo também pode ser importado por programas de análise estatística, como o R (figuras 26 e 27).

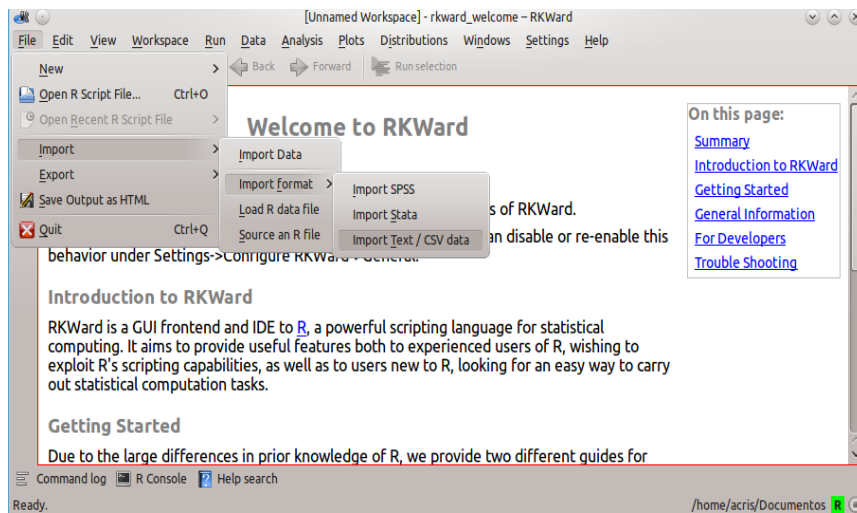


Ilustração 26: Exemplo de caminho para usar os dados da tabela CSV no programa de estatística R, pela interface do Rkward.

dadosSemiotica: projetos dS